Available online at www.sciencedirect.com
ScienceDirect
Journal of Approximation Theory 144 (2007) 84–102
Journal of Approximation Theory
www.elsevier.com/locate/jat

The asymptotic properties of the spectrum of nonsymmetrically perturbed Jacobi matrix sequences

Leonid Golinskii[a,*,1], Stefano Serra-Capizzano[b,2]

[a] *Mathematical Division, Institute for Low Temperature Physics, 47 Lenin ave, Kharkov 61103, Ukraine*
[b] *Department of Physics and Mathematics, University of "Insubria", Via Valleggio 11, 22100 Como, Italy*

Received 22 February 2006; received in revised form 2 May 2006; accepted 5 May 2006

Communicated by Sergey Krushchev
Available online 5 July 2006

Abstract

Under the mild trace-norm assumptions, we show that the eigenvalues of an arbitrary (non-Hermitian) complex perturbation of a Jacobi matrix sequence (not necessarily real) are still distributed as the real-valued function $2\cos t$ on $[0, \pi]$ which characterizes the nonperturbed case. In this way the real interval $[-2, 2]$ is still a cluster for the asymptotic joint spectrum and, moreover, $[-2, 2]$ still attracts strongly (with infinite order) the perturbed matrix sequence. The results follow in a straightforward way from more general facts that we prove in an asymptotic linear algebra framework and are plainly generalized to the case of matrix-valued symbols, which arises when dealing with orthogonal polynomials with asymptotically periodic recurrence coefficients.

MSC: 15A18; 15A12; 47B36; 47B65

Keywords: Matrix sequence; Joint eigenvalue distribution; Jacobi matrix; GLT sequence; Mergelyan Theorem

* Corresponding author.
E-mail addresses: golinskii@ilt.kharkov.ua (L. Golinskii), stefano.serrac@uninsubria.it (S. Serra-Capizzano).

[1] The work of the first author is supported in part by INTAS Research network 03-51-6637.

[2] The work of the second author is supported in part by MIUR Grant no. 2002014121.

doi:10.1016/j.jat.2006.05.002

1. Introduction and preliminary discussion

Consider the matrix J_n^0 of size n defined as

$$J_n^0 = \begin{bmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & 1 & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & 1 & 0 \end{bmatrix}. \tag{1}$$

The former matrix is the Toeplitz matrix $T_n(a)$ generated by $a(t) = 2\cos t$ in the following sense: given a Lebesgue integrable function b defined on $[-\pi, \pi)$ (and periodically extended on $\mathbb{R}$), the matrix $T_n(b)$ has order n and entries $(T_n(b))_{p,q} = \hat{b}_{p-q}$, $p, q = 1, \ldots, n$. Here, $\hat{b}_j$ is the jth Fourier coefficient of b, i.e.,

$$\hat{b}_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} b(t) \exp(-ijt)\, dt, \quad j \in \mathbb{Z}, \; i^2 = -1.$$

In the specific case (1) the eigenvalues are explicitly known, and they coincide with the evaluation of $a(t)$ on the uniform grid $j\pi/(n+1)$ on $[0, \pi]$. If J_n^0 is replaced by a more general Jacobi matrix

$$J_n = \begin{bmatrix} b_0 & a_1 & & & \\ a_1 & b_1 & a_2 & & \\ & a_2 & \ddots & \ddots & \\ & & \ddots & \ddots & a_{n-1} \\ & & & a_{n-1} & b_{n-1} \end{bmatrix}, \tag{2}$$

where $a_j \in \mathbb{R}$ tends to 1 and $b_j \in \mathbb{R}$ tends to 0 as $j \to \infty$, then its eigenvalues are no longer explicitly known, but they are again an approximation of the evaluation of $a(t)$ on the same grid. This result can be obtained directly from the GLT theory (see [16,17]), and more precisely, $\forall F \in \mathcal{C}_0(\mathbb{C})$ ($\mathcal{C}_0(\mathbb{C})$ is the set of all continuous functions having a bounded support), we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\lambda \in \Sigma_n} F(\lambda) = \frac{1}{\pi} \int_0^{\pi} F(2\cos t)\, dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(2\cos t)\, dt. \tag{3}$$

Here and in what follows, Σ_n stands for the collection of all eigenvalues of J_n counted with their multiplicity, the function $2\cos t$ is also called the symbol of $\{J_n\}$, and we write $\{J_n\} \sim_\lambda (2\cos t, [-\pi, \pi])$. In the orthogonal polynomials community this result is known, often under the unnecessary condition $a_j > 0$, in the form

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} F(x_{j,n}) = \frac{1}{\pi} \int_{-2}^{2} \frac{F(x)\, dx}{\sqrt{4 - x^2}},$$

that is the weak*-convergence of the counting measures of the zeros $\{x_{j,n}\}_{j=1}^n$ of orthonormal polynomials $\{p_n\}$ to the equilibrium measure of the support of the orthogonality measure (see, e.g., [13, Section 4.9; 18, Chapter 2]). Observe that the set of zeros $\{x_{j,n}\}_{j=1}^n$ is exactly the set Σ_n considered in the left-hand side of (3).

Let us set up the formal definitions. For any function F defined on $\mathbb{C}$ and any matrix A_n of size n, with the eigenvalues $\lambda_j(A_n)$, $j = 1, \ldots, n$, the symbol $\Sigma_\lambda(F, A_n)$ stands for the mean

$$\Sigma_\lambda(F, A_n) := \frac{1}{n} \sum_{j=1}^{n} F\left(\lambda_j(A_n)\right) = \frac{1}{n} \sum_{\lambda \in \Sigma_n} F(\lambda).$$

A generic sequence of matrices $\{A_n\} := \{A_n\}_n$ (A_n of size n) will be referred to as a *matrix sequence*.

Definition 1.1. A matrix sequence $\{A_n\}$ is *distributed* (*in the sense of the eigenvalues*) *as a measurable function* θ, defined on a set $G \subset \mathbb{R}^q$ of finite and positive Lebesgue measure $m(G)$, if $\forall F \in \mathcal{C}_0(\mathbb{C})$, the following limit relation holds:

$$\lim_{n\to\infty} \Sigma_\lambda(F, A_n) = \frac{1}{m(G)} \int_G F(\theta(t))\, dt. \tag{4}$$

In this case we write in short $\{A_n\} \sim_\lambda (\theta, G)$. Moreover, two sequences $\{A_n\}$ and $\{B_n\}$ are *equally distributed* if $\forall F \in \mathcal{C}_0(\mathbb{C})$, we have

$$\lim_{n\to\infty} [\Sigma_\lambda(F, B_n) - \Sigma_\lambda(F, A_n)] = 0. \tag{5}$$

Note that two sequences having the same distribution function are equally distributed. On the other hand, two equally distributed sequences do not need to have a distribution function. However, if one of them has a distribution function, then the other necessarily shares the same distribution: the derivation is immediate from the definitions (for an example see [15, Remark 6.1]).

Along with the distribution in the sense of eigenvalues (weak*-convergence) we will study another asymptotic property of the spectra Σ_n called here the *clustering*.

Definition 1.2. A matrix sequence $\{A_n\}$ is *properly* (*or strongly*) *clustered at* $s \in \mathbb{C}$ (the eigenvalue sense), if for any $\varepsilon > 0$ the number of the eigenvalues of A_n off the disk

$$D(s, \varepsilon) := \{z : |z - s| < \varepsilon\}$$

can be bounded by a pure constant q_ε possibly depending on ε, but not on n. In other words

$$q_\varepsilon(n, s) := \#\{\lambda_j(A_n) : \lambda_j \notin D(s, \varepsilon)\} = O(1), \quad n \to \infty.$$

If every A_n has only real eigenvalues (at least for all n large enough), then s is real and the disk $D(s, \varepsilon)$ reduces to the interval $(s - \varepsilon, s + \varepsilon)$. Furthermore, $\{A_n\}$ is *properly* (*or strongly*) *clustered at a nonempty closed set* $S \subset \mathbb{C}$ (*in the eigenvalue sense*) *if for any* $\varepsilon > 0$

$$q_\varepsilon(n, S) := \#\left\{\lambda_j(A_n) : \lambda_j \notin D(S, \varepsilon) := \bigcup_{s\in S} D(s, \varepsilon)\right\} = O(1), \quad n \to \infty, \tag{6}$$

$D(S, \varepsilon)$ is the ε-neighborhood of S, and if every A_n has only real eigenvalues, then S has to be a nonempty closed subset of $\mathbb{R}$. Finally, the term "properly (or strongly)" is replaced by "weakly", if

$$q_\varepsilon(n, s) = o(n) \quad \left(q_\varepsilon(n, S) = o(n)\right), \quad n \to \infty,$$

in the case of a point s (a closed set S), respectively.

It is clear that $\{A_n\} \sim_\lambda (\theta, G)$ with $\theta \equiv s$ a constant function is equivalent to $\{A_n\}$ being weakly clustered at $s \in \mathbb{C}$ (for more results and relations among the notions of equal distribution, equal localization, spectral distribution, spectral clustering etc., see [15, Section 4]).

We will primarily be interested in the special situation, when J_n are viewed as $n \times n$ principal blocks of an infinite Jacobi matrix J_∞ (background), P_∞ is a complex Jacobi matrix (perturbation), $A_\infty = J_\infty + P_\infty$ and $A_n = J_n + P_n$ are the $n \times n$ principal blocks of A_∞ (so A_{n+1} is the one step extension of A_n). In fact, the main results hold in much more general setting, when no relation between A_{n+1} and A_n is presumed.

The main conditions we impose on P_∞ are of two types.

(i) $\|P_n\|_1 = o(n)$ as $n \to \infty$, where $\|\cdot\|_1$ is the trace-norm of a matrix (i.e., the sum of its singular values, see [3]). This condition is equivalent to

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \big(|p_{j,j-1}| + |p_{j,j}| + |p_{j,j+1}|\big) = 0, \quad P_\infty = \{p_{j,k}\}_{j,k=1}^{\infty}. \tag{7}$$

The latter means the Cesàro convergence of the entries of P_∞ to zero. P_∞ is now called the *Cesàro compact Jacobi matrix* (cf. [6,7]).

(ii) $\|P_n\|_1 = O(1)$ as $n \to \infty$, that is,

$$\limsup_{n\to\infty} \sum_{j=1}^{n} \big(|p_{j,j-1}| + |p_{j,j}| + |p_{j,j+1}|\big) < \infty, \tag{8}$$

and so A_∞ is the trace class perturbation of J_∞.

We point out that the trace-norm is useful in the theoretical derivations, while the conditions on the entries are easy to check in practice. Moreover, the equivalence of the trace-norm and entry-wise conditions in (i) and (ii) is well known (cf. [10, Section 2]). Nevertheless, we give the proof in Appendix for two reasons: we deduce better equivalence constants and the proposed matrix-theoretic proof is new and elementary.

We proceed as follows. In Section 2 relevant relations between the notion of distribution in the sense of eigenvalues, clustering, and attracting properties of matrix sequences are discussed. Our main results are stated and proved in Section 3. In particular, Theorems 3.4 and 3.5 allow to study non-Hermitian perturbations of Hermitian matrix sequences. As a straightforward consequence we obtain the clustering for zeros of the system of polynomials satisfying the three-term recurrence relation with complex coefficients. Finally, in Section 4 we examine the case of block Toeplitz and asymptotically periodic Jacobi matrix sequences, and then we discuss further extensions and generalizations.

2. Clustering and attracting

Let us recall the notion of the essential range which plays an important role in the study of asymptotic properties of the spectrum.

Definition 2.1. Given a measurable complex-valued function θ defined on a Lebesgue measurable set G, the *essential range of* θ is the set $S(\theta)$ of points $s \in \mathbb{C}$ such that, for every $\varepsilon > 0$, the Lebesgue measure of the set $\theta^{(-1)}(D(s,\varepsilon)) := \{t \in G : \theta(t) \in D(s,\varepsilon)\}$ is positive. The function

θ is *essentially bounded* if its essential range is bounded. Finally, if θ is real-valued, then the essential supremum (infimum) is defined as the supremum (infimum) of its essential range.

$S(\theta)$ is clearly a closed set (its complement is open), and moreover

$$S(\theta) = \bigcap \{B - \text{closed set} : m(\theta^{(-1)}(B)) = m(G)\},$$

where $m(X)$ is the Lebesgue measure of a set X.

In the case of a sequence $\{A_n\}$, bounded in the operator norm, a further mathematical instrument that we need is a way for relating formula (4), with F being an arbitrary polynomial, to the same formula in its full extent, i.e., with F being a continuous function. The answer is partly contained in the Mergelyan Theorem and not completely positive. We need assumptions on the essential range of the symbol θ and a priori assumptions on the clustering properties of the sequence $\{A_n\}$. The reason is in part due to the barrier given by the Mergelyan Theorem stating that the closure in the uniform norm of the polynomials on a compact set S is given by the set of all continuous functions on S which are holomorphic in its interior, provided that $\mathbb{C}\backslash S$ is connected (for the proof see [14, Theorem 20.5, pp. 423–427]). Therefore, the polynomial space is able to approximate every continuous function on S if and only if S has empty interior in $\mathbb{C}$ and $\mathbb{C}\backslash S$ is connected.

Theorem 2.2. *Assume that a matrix sequence $\{A_n\}$ is weakly clustered at a compact set $S \subset \mathbb{C}$, that $\mathbb{C}\backslash S$ is connected, and that the spectra Σ_n are uniformly bounded, i.e., $|\lambda| < C$, $\lambda \in \Sigma_n$, for all n. Assume further that* (4) *holds with F being any polynomial of an arbitrary fixed degree, and the essential range of θ is contained in S. Then relation* (4) *is true for every continuous function F with a bounded support which is holomorphic in the interior of S. Moreover, if the interior of S is empty, then $\{A_n\}$ is distributed as θ on its domain G.*

Proof. In the argument we follow Tilli (see [20], the proof of Theorem 3). Take F continuous over S and holomorphic in its interior. By the Mergelyan Theorem, for every $\varepsilon > 0$, we can find a polynomial p such that $|p(z) - F(z)| \leqslant \varepsilon$ for every $z \in S$. Since the essential range of θ is contained in S, it is clear that $|p(\theta(t)) - F(\theta(t))| \leqslant \varepsilon$ a.e. in its domain G. Therefore,

$$\left| \frac{1}{m(G)} \int_G F(\theta(t))\, dt - \frac{1}{m(G)} \int_G p(\theta(t))\, dt \right| \leqslant \frac{\varepsilon}{m(G)} \int_G dt = \varepsilon. \tag{9}$$

Next, we treat the left-hand side of (4). By the definition of clustering, for any fixed $\varepsilon' > 0$, we have

$$\#\{\lambda \in \Sigma_n,\ |\lambda - z| \geqslant \varepsilon',\ \forall z \in S\} = \#\{\lambda \in \Sigma_n,\ \lambda \notin D(S, \varepsilon')\} = o(n).$$

Moreover, by the hypothesis of the uniform boundedness of Σ_n, $|\lambda| < C$ for every $\lambda \in \Sigma_n$ with a pure constant C independent of n. Therefore, by extending F outside S in such a way that it is continuous with a bounded support, we infer

$$\left| \frac{1}{n} \sum_{\lambda \in \Sigma_n,\ \lambda \notin D(S,\varepsilon')} F(\lambda) \right| \leqslant \frac{M}{n} \#\{\lambda \in \Sigma_n,\ \lambda \notin D(S, \varepsilon')\} = o(1),$$

$$\left| \frac{1}{n} \sum_{\lambda \in \Sigma_n,\ \lambda \notin D(S,\varepsilon')} p(\lambda) \right| \leqslant \frac{M}{n} \#\{\lambda \in \Sigma_n,\ \lambda \notin D(S, \varepsilon')\} = o(1),$$

with $M = \max(\|F\|_\infty, \|p\|_\infty)$, and the infinity norms are taken over $\{z \in \mathbb{C},\ |z| \leqslant C\}$. Consequently, by setting $\Delta = |\Sigma_\lambda(F - p, A_n)|$, we deduce

$$\begin{aligned}
\Delta &= \left| \frac{1}{n} \sum_{\lambda \in \Sigma_n} (F(\lambda) - p(\lambda)) \right| \leqslant \frac{1}{n} \sum_{\lambda \in \Sigma_n} |F(\lambda) - p(\lambda)| \\
&= \frac{1}{n} \sum_{\lambda \in \Sigma_n,\ \lambda \in D(S,\varepsilon')} |F(\lambda) - p(\lambda)| + \frac{1}{n} \sum_{\lambda \in \Sigma_n,\ \lambda \notin D(S,\varepsilon')} |F(\lambda) - p(\lambda)| \\
&\leqslant \frac{1}{n} \sum_{\lambda \in \Sigma_n,\ \lambda \in D(S,\varepsilon')} |F(\lambda) - p(\lambda)| + o(1) \\
&= \frac{1}{n} \sum_{\lambda \in \Sigma_n,\ \lambda \in S} |F(\lambda) - p(\lambda)| + \frac{1}{n} \sum_{\lambda \in \Sigma_n,\ \lambda \in D(S,\varepsilon') \setminus S} |F(\lambda) - p(\lambda)| + o(1).
\end{aligned}$$

For $\lambda \in S$ we use $|F(\lambda) - p(\lambda)| \leqslant \varepsilon$, and, for $\lambda \in D(S, \varepsilon') \setminus S$, we write

$$|F(\lambda) - p(\lambda)| \leqslant |F(\lambda) - F(\lambda')| + |F(\lambda') - p(\lambda')| + |p(\lambda') - p(\lambda)|, \quad |\lambda - \lambda'| < \varepsilon', \quad \lambda' \in S,$$

so that $|F(\lambda) - p(\lambda)| \leqslant c_1(\varepsilon') + \varepsilon + c_2(\varepsilon, \varepsilon') \equiv \theta(\varepsilon, \varepsilon')$ with

$$\lim_{\varepsilon \to 0} \lim_{\varepsilon' \to 0} \theta(\varepsilon, \varepsilon') = 0. \tag{10}$$

Hence

$$\Delta \leqslant \varepsilon + \theta(\varepsilon, \varepsilon') + o(1). \tag{11}$$

Furthermore, from the hypothesis of the theorem we have

$$\lim_{n \to \infty} \Sigma_\lambda(p, A_n) = \frac{1}{m(G)} \int_G p(\theta(t))\, dt. \tag{12}$$

Since ε and ε' are arbitrary, it is clear that relations (9)–(12) imply (4) to hold for F as well. Finally, when S has empty interior, we have no restriction on F except for being continuous with a bounded support, and therefore what we have proved is equivalent to $\{A_n\} \sim_\lambda (\theta, G)$. $\square$

To proceed further, we need a notion which is essential in the orthogonal polynomials theory.

Definition 2.3. A matrix sequence $\{A_n\}$ is *strongly attracted by* $s \in \mathbb{C}$ if

$$\lim_{n \to \infty} \operatorname{dist}(s, \Sigma_n) = 0, \tag{13}$$

where $\operatorname{dist}(X, Y)$ is the usual Euclidean distance between two subsets X and Y of the complex plane. Furthermore, let us order the eigenvalues according to its distance from s, i.e.,

$$|\lambda_1(A_n) - s| \leqslant |\lambda_2(A_n) - s| \leqslant \cdots \leqslant |\lambda_n(A_n) - s|.$$

We say that the attraction is of order $r(s) \in \mathbb{N}$, $r(s) \geqslant 1$ is a fixed number, if

$$\lim_{n \to \infty} |\lambda_{r(s)}(A_n) - s| = 0, \quad \liminf_{n \to \infty} |\lambda_{r(s)+1}(A_n) - s| > 0.$$

The attraction is of order $r(s) = \infty$ if

$$\lim_{n\to\infty} |\lambda_j(A_n) - s| = 0$$

for every fixed j. Finally, the term "strong or strongly" is replaced by "weak or weakly" if lim is replaced by lim inf in (13).

It is not hard to ascertain, that if $\{A_n\}$ is at least weakly clustered at a point s, then s strongly attracts $\{A_n\}$ with infinite order. Indeed, s is an attracting point of finite order implies

$$\lim_{n\to\infty} \frac{\#\{\lambda \in \Sigma_n : \lambda \notin D(s,\delta)\}}{n} = 1$$

for some $\delta > 0$, that is impossible in the case when $\{A_n\}$ is weakly clustered at s. On the other hand, there are sequences which are strongly attracted by s with infinite order, but not even weakly clustered at s.

The notions previously introduced in this section are intimately related, as emphasized in the following theorem.

Theorem 2.4. *Let θ be a measurable function defined on G with finite and positive Lebesgue measure, and $S = S(\theta)$ be the essential range of θ. Let $\{A_n\}$ be a matrix sequence distributed as θ in the sense of eigenvalues. Then*

(a) *$S(\theta)$ is a weak cluster for $\{A_n\}$;*
(b) *each point $s \in S(\theta)$ strongly attracts Σ_n with infinite order $r(s) = \infty$.*

Proof. (a) Given $\varepsilon > 0$, we apply (4) with the test function F_ε of the form

$$F_\varepsilon(z) = \begin{cases} 1 & \text{for } z \in D(S,\varepsilon/2) \cap D(0,1/\varepsilon), \\ 0 & \text{for } z \in \mathbb{C}\setminus (D(S,\varepsilon) \cap D(0,2/\varepsilon)), \end{cases} \qquad 0 \leqslant F_\varepsilon \leqslant 1.$$

It is clear that

$$\begin{aligned} \Sigma_\lambda(F_\varepsilon, A_n) &\leqslant \frac{\#\{\lambda \in \Sigma_n : \lambda \in (D(S,\varepsilon) \cap D(0,2/\varepsilon))\}}{n} \\ &\leqslant \frac{\#\{\lambda \in \Sigma_n : \lambda \in D(S,\varepsilon)\}}{n} = 1 - \frac{q_\varepsilon(n,S)}{n}, \end{aligned}$$

$q_\varepsilon(n,S)$ is defined in (6), and hence

$$\liminf_{n\to\infty} \Sigma_\lambda(F_\varepsilon, A_n) \leqslant 1 - \limsup_{n\to\infty} \frac{q_\varepsilon(n,S)}{n}.$$

By the assumption there exists

$$\lim_{n\to\infty} \Sigma_\lambda(F_\varepsilon, A_n) = \frac{1}{m(G)} \int_G F(\theta(t))\,dt \geqslant \frac{m\{\theta^{(-1)}(D(S,\varepsilon/2) \cap D(0,1/\varepsilon))\}}{m(G)}.$$

We have

$$\theta^{(-1)}(D(S,\varepsilon/2) \cap D(0,1/\varepsilon)) = \theta^{(-1)}(D(S,\varepsilon/2)) \cap \theta^{(-1)}(D(0,1/\varepsilon)) = \Gamma_\varepsilon \cap \Delta_\varepsilon$$

and hence

$$1 - \limsup_{n\to\infty} \frac{q_\varepsilon(n, S)}{n} \geqslant \frac{m(\Gamma_\varepsilon \cap \Delta_\varepsilon)}{m(G)}. \tag{14}$$

By the definition of the essential range, the right-hand side in (14) tends to 1 as $\varepsilon \to 0$, and therefore $\lim_{n\to\infty} n^{-1} q_\varepsilon(n, S) = 0$, as needed.

(b) Let $s \in S$ and $\varepsilon > 0$. Construct F_ε by

$$F_\varepsilon(z) = \begin{cases} 1 & \text{for } z \in D(s, \varepsilon), \\ 0 & \text{for } z \in \mathbb{C}\backslash(D(s, 2\varepsilon)), \end{cases} \quad 0 \leqslant F_\varepsilon \leqslant 1.$$

Since F_ε is dominated by the characteristic function of $D(s, 2\varepsilon)$, we see that

$$\frac{\#\{\lambda \in \Sigma_n : \lambda \in D(s, 2\varepsilon)\}}{n} \geqslant \Sigma_\lambda(F_\varepsilon, A_n).$$

But $\{A_n\} \sim_\lambda (\theta, G)$. As a consequence, by employing F_ε as the test function, we obtain

$$\lim_{n\to\infty} \Sigma_\lambda(F_\varepsilon, A_n) = \frac{1}{m(G)} \int_G F_\varepsilon(\theta(t))\, dt \geqslant \frac{m\{\theta^{(-1)}(D(s, \varepsilon))\}}{m(G)},$$

since F_ε dominates the characteristic function of $D(s, \varepsilon)$. By the definition of the essential range, the right-hand side is strictly positive and hence

$$\liminf_{n\to\infty} \frac{\#\{\lambda \in \Sigma_n : \lambda \in D(s, 2\varepsilon)\}}{n} > 0.$$

The latter means exactly that s attracts Σ_n with order $r(s) = \infty$, and the proof is concluded. □

The final result of this section demonstrates the stability of the clustering under certain perturbations (cf. [15, Corollary 4.1]).

Proposition 2.5. *Let $\{X_n\}$ and $\{Y_n\}$ be two Hermitian matrix sequences, let M be a closed subset of the real line, and assume that $\|X_n - Y_n\|_1 = o(n)$ $\big(\|X_n - Y_n\|_1 = O(1)\big)$. Then $\{X_n\}$ is weakly (strongly) clustered at M if and only if the same property holds for $\{Y_n\}$.*

Proof. Let $\lambda_j(X_n), \lambda_j(Y_n)$ be the eigenvalues of X_n and Y_n, respectively, labelled in the decreasing order. For an arbitrary $\varepsilon > 0$ we introduce three sets of indices

$$I(X_n, \varepsilon) = \{j = 1, 2, \ldots, n : \operatorname{dist}(\lambda_j(X_n), M) > \varepsilon\},$$
$$I(Y_n, \varepsilon) = \{j = 1, 2, \ldots, n : \operatorname{dist}(\lambda_j(Y_n), M) > \varepsilon\},$$
$$I(X_n, Y_n, \varepsilon) = \{j = 1, 2, \ldots, n : |\lambda_j(X_n) - \lambda_j(Y_n)| > \varepsilon\}.$$

Let us denote by $|I(X_n, \varepsilon)|$, $|I(Y_n, \varepsilon)|$, and $|I(X_n, Y_n, \varepsilon)|$ their cardinalities. It is clear that

$$I(X_n, \varepsilon) \subset I\left(X_n, Y_n, \frac{\varepsilon}{2}\right) \cup I\left(Y_n, \frac{\varepsilon}{2}\right).$$

Thus

$$|I(X_n, \varepsilon)| \leqslant \left|I\left(X_n, Y_n, \frac{\varepsilon}{2}\right)\right| + \left|I\left(Y_n, \frac{\varepsilon}{2}\right)\right|.$$

According to the Lidskii–Mirsky–Wielandt Theorem (see [3, Theorem IV.3.4 and Example IV.3.5])

$$\sum_{j=1}^{n} \left|\lambda_j(X_n) - \lambda_j(Y_n)\right| \leqslant \|X_n - Y_n\|_1,$$

from which

$$\varepsilon\ |I\ (X_n, Y_n, \varepsilon)| \leqslant \sum_{j \in I(X_n, Y_n, \varepsilon)} \left|\lambda_j(X_n) - \lambda_j(Y_n)\right| \leqslant \|X_n - Y_n\|_1.$$

Hence

$$|I(X_n, \varepsilon)| \leqslant \frac{2}{\varepsilon}\ \|X_n - Y_n\|_1 + \left|I\left(Y_n, \frac{\varepsilon}{2}\right)\right|.$$

The rest is plain. □

3. Non-Hermitian perturbations of Hermitian matrix sequences

First, we recall the definition of real and imaginary parts of a matrix. Given a square matrix A, we define $\mathrm{Re}(A)$ and $\mathrm{Im}(A)$ as $(A + A^*)/2$ and $(A - A^*)/(2i)$, respectively, where X^* denotes the conjugate transpose of the matrix X. In this way, in analogy to the complex field, we naturally have $A = \mathrm{Re}(A) + i\ \mathrm{Im}(A)$.

The result below is the well-known Ky Fan–Mirski Theorem (see, e.g., [3, Proposition III. 5.3]).

Theorem 3.1. *Let A be a square matrix of size n, and $\lambda_j(A)$, $\lambda_j(\mathrm{Im}(A))$, $j = 1, \ldots, n$ the eigenvalues of A and $\mathrm{Im}(A)$, respectively, labelled in the decreasing order, so that $\mathrm{Im}(\lambda_1(A)) \geqslant \mathrm{Im}(\lambda_2(A)) \geqslant \cdots \geqslant \mathrm{Im}(\lambda_n(A))$ and $\lambda_1(\mathrm{Im}(A)) \geqslant \lambda_2(\mathrm{Im}(A)) \geqslant \cdots \geqslant \lambda_n(\mathrm{Im}(A))$. Then*

$$\sum_{j=1}^{q} \mathrm{Im}(\lambda_j(A)) \leqslant \sum_{j=1}^{q} \lambda_j(\mathrm{Im}(A)), \qquad q = 1, \ldots, n, \tag{15}$$

and the equality prevails for $q = n$.

Equivalently, let $\lambda_j(A)$ and $\lambda_j(\mathrm{Re}(A))$, $j = 1, \ldots, n$, be the eigenvalues of A and $\mathrm{Re}(A)$, respectively, labelled in the decreasing order, so that $\mathrm{Re}(\lambda_1(A)) \geqslant \mathrm{Re}(\lambda_2(A)) \geqslant \cdots \geqslant \mathrm{Re}(\lambda_n(A))$ and $\lambda_1(\mathrm{Re}(A)) \geqslant \lambda_2(\mathrm{Re}(A)) \geqslant \cdots \geqslant \lambda_n(\mathrm{Re}(A))$. Then

$$\sum_{j=1}^{q} \mathrm{Re}(\lambda_j(A)) \leqslant \sum_{j=1}^{q} \lambda_j(\mathrm{Re}(A)), \qquad q = 1, \ldots, n, \tag{16}$$

and the equality prevails for $q = n$.

The next statement provides a simple bound for the number of essentially nonreal eigenvalues of a matrix A. In what follows $\Sigma(X)$ always stands for the set of all eigenvalues of a matrix X: $\Sigma(X) = \{\lambda_j(X)\}_{j=1}^n$.

Lemma 3.2. *Let $A = \mathrm{Re}(A) + i\ \mathrm{Im}(A)$. Then for an arbitrary $\varepsilon > 0$*

$$\#\{\lambda \in \Sigma(A) : |\mathrm{Im}(\lambda)| > \varepsilon\} \leqslant \frac{\|\mathrm{Im}(A)\|_1}{\varepsilon}. \tag{17}$$

Moreover, if for some real c, d we have $c \leqslant \lambda_j(\mathrm{Re}(A)) \leqslant d$ *for all j, then* $c \leqslant \mathrm{Re}(\lambda_j(A)) \leqslant d$ *and*

$$\#\{\lambda \in \Sigma(A) : \lambda \notin D([c,d],\varepsilon)\} \leqslant \frac{\|\mathrm{Im}(A)\|_1}{\varepsilon}. \tag{18}$$

Proof. Denote by

$$m^+ := \sum_{\lambda\in\Sigma(\mathrm{Im}(A)),\ \lambda\geqslant 0} \lambda, \quad \left(m^- := \sum_{\lambda\in\Sigma(\mathrm{Im}(A)),\ \lambda<0} |\lambda|\right)$$

the positive (negative) mass of the eigenvalues of $\mathrm{Im}(A)$. Since $\mathrm{Im}(A)$ is Hermitian, its trace-norm equals the sum of the absolute values of its eigenvalues, so $\|\mathrm{Im}(A)\|_1 = m^+ + m^-$. We apply the first part of Theorem 3.1 for A and $-A$ to obtain

$$r^+ := \sum_{\lambda\in\Sigma(A),\ \mathrm{Im}(\lambda)\geqslant 0} \mathrm{Im}(\lambda) \leqslant m^+, \quad r^- := \sum_{\lambda\in\Sigma(A),\,\mathrm{Im}(\lambda)<0} |\mathrm{Im}(\lambda)| \leqslant m^-. \tag{19}$$

Therefore, if we take an arbitrary $\varepsilon > 0$, the number of the eigenvalues of A whose imaginary part is bigger than ε has to be bounded by $\|\mathrm{Im}(A)\|_1/\varepsilon$. Indeed,

$$\begin{aligned}\|\mathrm{Im}(A)\|_1 &= m^+ + m^- \geqslant r^+ + r^- = \sum_{\lambda\in\Sigma(A)} |\mathrm{Im}(\lambda)| \geqslant \sum_{\lambda\in\Sigma(A),\ |\mathrm{Im}(\lambda)|>\varepsilon} |\mathrm{Im}(\lambda)| \\ &\geqslant \sum_{\lambda\in\Sigma(A),\ |\mathrm{Im}(\lambda)|>\varepsilon} \varepsilon = \varepsilon \cdot \#\{\lambda \in \Sigma(A),\ |\mathrm{Im}(\lambda)| > \varepsilon\},\end{aligned}$$

as needed.

Next, let λ be an eigenvalue of A corresponding to an eigenvector $\mathbf{x}$. Then

$$\lambda = \frac{\mathbf{x}^* A\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \frac{\mathbf{x}^*\mathrm{Re}(A)\mathbf{x}}{\mathbf{x}^*\mathbf{x}} + i\,\frac{\mathbf{x}^*\,\mathrm{Im}(A)\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$$

which implies that $\mathrm{Re}(\lambda) \in [c,d]$, since, by the assumption, every eigenvalue of $\mathrm{Re}(A)$ belongs to $[c,d]$. Therefore (18) follows from (17). □

Corollary 3.3. *Let* $\{A_n\}$ *be a matrix sequence such that* $\|\mathrm{Im}(A_n)\|_1 = o(n)$ *as* $n \to \infty$. *Then* $q_\varepsilon(n,\mathbb{R}) = o(n)$, *so* $\{A_n\}$ *is weakly clustered at* $\mathbb{R}$. *Moreover, if all the eigenvalues of* $\mathrm{Re}(A_n)$ *are in* $[c,d]$, *then all the eigenvalues of* A_n *have real parts in the same interval and* $q_\varepsilon(n,[c,d]) = o(n)$. *The same result holds if* $o(n)$ *is replaced by* $O(1)$ *and "weakly clustered" by "strongly clustered".*

The following result establishes a link between distributions of the Hermitian sequence $\{\mathrm{Re}(A_n)\}$ and the sequence $\{A_n\}$. As a matter of fact, we will prove a more general statement concerning non-Hermitian perturbations of Hermitian matrix sequences. As usual, $\|X\|$ stands for the operator (spectral) norm of a matrix X.

Theorem 3.4. *Let* $\{B_n\}$ *and* $\{C_n\}$ *be two matrix sequences, where* B_n *is Hermitian and* $A_n = B_n + C_n$. *Assume further that* $\{B_n\}$ *is distributed as* (θ, G), *G of finite and positive Lebesgue measure, both* $\|B_n\|$ *and* $\|C_n\|$ *are uniformly bounded by a positive constant C independent of n, and* $\|C_n\|_1 = o(n)$, $n \to \infty$. *Then* θ *is real valued and* $\{A_n\}$ *is distributed as* (θ, G) *in the sense of the eigenvalues. In particular, if* $S(\theta)$ *is the essential range of* θ, *then* $\{A_n\}$ *is weakly clustered at* $S(\theta)$, *and* $S(\theta)$ *strongly attracts the spectra of* $\{A_n\}$ *with infinite order of attraction for any of its points.*

Proof. Denote by tr X the trace of a matrix X, that is, the sum of its diagonal entries (or the sum of its eigenvalues)

$$\operatorname{tr} X = \sum_{\lambda\in\Sigma(X)} \lambda = \sum_{k=1}^{n} (X)_{k,k},$$

so $\operatorname{tr} A_n - \operatorname{tr} B_n = \operatorname{tr} C_n$. As $|\operatorname{tr} X| \leqslant \|X\|_1$, the assumption on the trace-norm of C_n yields

$$\frac{1}{n}\sum_{\lambda\in\Sigma(A_n)} \lambda = \frac{1}{n}\sum_{\lambda\in\Sigma(B_n)} \lambda + o(1).$$

The latter is closely related to (4) with $F(z) = z$ (defined over the whole $\mathbb{C}$). Since $\{B_n\}$ is distributed as θ over G, we infer by (4)

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\lambda\in\Sigma(A_n)} \lambda = \lim_{n\to\infty}\frac{1}{n}\sum_{\lambda\in\Sigma(B_n)} \lambda = \frac{1}{m(G)}\int_G F(\theta(t))\,dt, \qquad F(z) = z, \tag{20}$$

where we are allowed to take $F(z) = z$ (which has an unbounded support), since by the premises of the theorem $\|A_n\| \leqslant 2C$ for all n, and then the spectra of $\{A_n\}$, $\{B_n\}$, and $\{C_n\}$ are all contained in the closed disk $\{|z| \leqslant 2C\}$. Equalities (20) can be viewed as the first step for proving a distribution relation for $\{A_n\}$, starting from same distribution relation for the Hermitian sequence $\{B_n\}$. The next step is to extend (20) to the case when F is an arbitrary polynomial of a fixed degree. By the linearity it suffices to consider only monomials. Clearly, for any fixed nonnegative integer q, the matrix A_n^q can be written as $A_n^q = B_n^q + R_{n,q}$ and, thanks to the Hölder-type inequalities for the Schatten p norms $\|XY\|_1 \leqslant \|X\| \cdot \|Y\|_1$ (see [3, Corollary IV.2.6]) we have $\|R_{n,q}\|_1 = o(n)$ as $n \to \infty$. Therefore, by repeating the same reasoning as above we deduce

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\lambda\in\Sigma(A_n)} \lambda^q = \lim_{n\to\infty}\frac{1}{n}\sum_{\lambda\in\Sigma(B_n)} \lambda^q = \frac{1}{m(G)}\int_G F(\theta(t))\,dt, \qquad F(z) = z^q. \tag{21}$$

To go over in (21) from polynomials to arbitrary continuous functions with bounded support we would like to invoke Theorem 2.2. So let us make sure that the rest of its hypothesis is satisfied. As we have already mentioned, $\|A_n\| \leqslant 2C$ for all n. Next, it is clear that

$$\|\mathrm{Re}(C_n)\|_1 \leqslant \|C_n\|_1 = o(n), \qquad \|\mathrm{Im}(C_n)\|_1 \leqslant \|C_n\|_1 = o(n) \tag{22}$$

as $n \to \infty$. Write $A_n = B_n + \mathrm{Re}(C_n) + i\,\mathrm{Im}(C_n)$. By Theorem 2.4 $\{B_n\}$ is weakly clustered at $S(\theta)$, and so is $\{\mathrm{Re}(A_n) = B_n + \mathrm{Re}(C_n)\}$ by Proposition 2.5. Note that $S(\theta)$ is now a compact set which lies in the interval $[-2C, 2C]$, and all the eigenvalues of $\mathrm{Re}(A_n)$ are in the same interval. Corollary 3.3 now claims that $\{A_n\}$ is weakly clustered at $[-2C, 2C] \supset S(\theta)$, and the application of Theorem 2.2 completes the proof. $\square$

The following theorem deals with the case of the strong clustering.

Theorem 3.5. *Let $\{B_n\}$ and $\{C_n\}$ be two matrix sequences, where B_n is Hermitian and $A_n = B_n + C_n$. Assume that $\{B_n\}$ is strongly clustered at $[c, d]$, $\|C_n\|_1 = O(1)$, $n \to \infty$ and $\|A_n\|$ is uniformly bounded by a positive constant C independent of n. Then $\{A_n\}$ is strongly clustered at $[c, d]$.*

Proof. Since now $\|\mathrm{Re}(C_n)\|_1 = O(1)$ and $\|\mathrm{Im}(C_n)\|_1 = O(1)$, both the related sequences are strongly clustered at zero by Proposition 2.5. A repeated application of the same proposition shows that $\{B_n + \mathrm{Re}(C_n)\}$ is strongly clustered at $[c, d]$. Although we are not allowed to invoke Corollary 3.3 at this point, since the eigenvalues of $\mathrm{Re}(A_n) = B_n + \mathrm{Re}(C_n)$ are not necessarily in $[c, d]$, we can follow a direct approach stemming from Theorem 3.1.

Since $\|A_n\| \leqslant C$, the real part of any eigenvalue of A_n belongs to $[-C, C]$ and the same is true for any eigenvalue of $\mathrm{Re}(A_n)$. For $\varepsilon > 0$, let $q_n^-(\varepsilon)$ be the number of eigenvalues of A_n whose real parts are below $c - \varepsilon$, and analogously, let $q_n^+(\varepsilon)$ be the number of eigenvalues of X_n whose real parts exceed $d + \varepsilon$. We want to prove that both $q_n^-(\varepsilon)$ and $q_n^+(\varepsilon)$ can be bounded by a constant possibly depending on ε, but independent of n. By (16) we have

$$\sum_{j=1}^{q_n^+(\varepsilon)} \mathrm{Re}(\lambda_j(A_n)) \leqslant \sum_{j=1}^{q_n^+(\varepsilon)} \lambda_j(\mathrm{Re}(A_n))$$

with

$$\mathrm{Re}(\lambda_1(A_n)) \geqslant \mathrm{Re}(\lambda_2(A_n)) \geqslant \cdots \geqslant \mathrm{Re}(\lambda_{q_n^+(\varepsilon)}(A_n)) > d + \varepsilon \geqslant \mathrm{Re}(\lambda_{q_n^+(\varepsilon)+1}(A_n)).$$

Therefore,

$$(d + \varepsilon) q_n^+(\varepsilon) \leqslant \sum_{j=1}^{q_n^+(\varepsilon)} \lambda_j(\mathrm{Re}(A_n)). \tag{23}$$

Thanks to the strong clustering of $\mathrm{Re}(A_n) = B_n + \mathrm{Re}(C_n)$, for every $\varepsilon' > 0$ there exists a positive constant $K(\varepsilon')$ independent of n such that the number of eigenvalues of $\mathrm{Re}(A_n)$ not belonging to $(c - \varepsilon', d + \varepsilon')$ is bounded by $K(\varepsilon')$. Consequently, we infer

$$\sum_{j=1}^{q_n^+(\varepsilon)} \lambda_j(\mathrm{Re}(A_n)) \leqslant C K(\varepsilon') + (d + \varepsilon')(q_n^+(\varepsilon) - K(\varepsilon'))^+ \tag{24}$$

with $(x)^+ = (x + |x|)/2$. Putting together (23) and (24), by choosing $\varepsilon' = \varepsilon/2$, we finally deduce

$$q_n^+(\varepsilon) \leqslant \frac{2CK(\varepsilon/2)}{\varepsilon},$$

where as requested, the right-hand side is independent of n. A similar reasoning on $-X_n$ gives the same bound on $q_n^-(\varepsilon)$, as claimed.

As for the imaginary parts of the eigenvalues of A_n, we can apply directly (17). As a consequence, the proof is complete. □

The latter result can be extended to the case of clustering at several intervals, the situation we will encounter later in Theorem 4.6.

Theorem 3.6. *Let $\{B_n\}$ and $\{C_n\}$ be two matrix sequences, where B_n is Hermitian and $A_n = B_n + C_n$. Let E be a union of m disjoint closed intervals (possibly, degenerate). Assume that $\{B_n\}$ is strongly clustered at E, $\|C_n\|_1 = O(1)$, $n \to \infty$ and $\|A_n\|$ is uniformly bounded by a positive constant C independent of n. Then $\{A_n\}$ is strongly clustered at E.*

Proof. We reduce this statement to the previous one. Denote

$$E = \bigcup_{j=1}^{m} [a_j, b_j], \quad a_1 \leqslant b_1 < a_2 \leqslant b_2 < \cdots < a_m \leqslant b_m,$$

and put $T(z) = \prod_{j=1}^{m} (z - a_j)(z - b_j)$. Obviously, $T(E) \in [\omega, 0]$, $\omega = \min_x T(x) < 0$, and $T(x) > 0$ for $x \in \mathbb{R} \backslash E$. By the Spectral Mapping Theorem (see e.g., [3, p. 5]) E is a strong cluster for $\{B_n\}$ yields $[\omega, 0]$ is a strong cluster for $\{T(B_n)\}$. Next, by the hypothesis of the theorem and the Hölder-type inequalities for the trace-norm

$$T(A_n) = T(B_n) + R_n, \quad \|R_n\|_1 = O(1), \quad n \to \infty.$$

We have the right to apply Theorem 3.5 to the matrix sequences $\{T(A_n)\}$, $\{T(B_n)\}$ to conclude that $\{T(A_n)\}$ is strongly clustered at $[\omega, 0]$. The repeated application of the Spectral Mapping Theorem completes the proof. □

Let us go back to the Jacobi matrix sequences described in the Introduction. Let

$$A_\infty = \begin{bmatrix} b_0 & c_1 & & & \\ a_1 & b_1 & c_2 & & \\ & a_2 & b_2 & c_3 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

be an infinite complex nonsymmetric Jacobi matrix with the bounded entries

$$\sup_n (|a_n| + |b_n| + |c_n|) \leqslant C < \infty. \tag{25}$$

As a simple consequence of Theorem 3.4, we can prove the following

Corollary 3.7. *Let A_∞ be the Cesàro compact perturbation of J^0_∞, that is,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (|1 - a_j| + |b_j| + |1 - c_j|) = 0, \tag{26}$$

and $\{A_n\}$ its principal $n \times n$ blocks. Then $\{A_n\}$ is distributed as $(2\cos t, [-\pi, \pi])$ in the sense of eigenvalues, weakly clustered at $[-2, 2]$, and $[-2, 2]$ strongly attracts the spectra of $\{A_n\}$ with infinite order of attraction for any of its points.

Proof. We apply Theorem 3.4 with $B_n = J^0_n$, $C_n = A_n - J^0_n$. $\{B_n\}$ is clearly distributed as $(2\cos t, [-\pi, \pi])$. Inequality (25) provides the uniform boundedness of $\|A_n\|$ and $\|C_n\|$. Finally, (26) is equivalent to $\|C_n\|_1 = o(n)$, and the result follows. □

Corollary 3.8. *Let A_∞ be trace class perturbation of J_∞, that is,*

$$\limsup_{n\to\infty} \sum_{j=1}^{n} (|1 - a_j| + |b_j| + |1 - c_j|) < \infty. \tag{27}$$

Then $\{A_n\}$ is distributed as $(2\cos t, [-\pi, \pi])$ in the sense of eigenvalues, strongly clustered at $[-2, 2]$, and $[-2, 2]$ strongly attracts the spectra of $\{A_n\}$ with infinite order of attraction for any of its points.

Proof. The only point to be proved is that the weak cluster is also strong, and this is implied by Theorem 3.5. □

If we are concerned only about the clustering of the spectrum, the more elementary Corollary 3.3 does the job. In this case the assumption $\mathrm{Im}(A_\infty)$ being a Cesàro compact perturbation of J^0_∞, that is,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (|\mathrm{Im}(b_j)| + |a_j - \bar{c}_j|) = 0,$$

already guarantees that $\{A_n\}$ is weakly clustered at $\mathbb{R}$. Moreover, if all the eigenvalues of $\mathrm{Re}(A_n)$ are in $[c, d]$, then $\{A_n\}$ is weakly clustered at $[c, d]$.

It is worth pointing out that $\Sigma(A_n)$ now agrees with the set of all zeros of the polynomial p_n which satisfies the three-term recurrence relation

$$zp_j(z) = a_j p_{j-1}(z) + b_j p_j(z) + c_{j+1} p_{j+1}(z), \quad j \in \mathbb{Z}_+ \tag{28}$$

$p_{-1} = 0$, $p_0 = 1$. Such polynomials are studied systematically in the theory of Padé approximations and continued *J*-fractions. More precisely, p_n is the denominator of the nth diagonal Padé approximant and its zeros are the poles of this Padé approximant. In turn, the closed interval $[-2, 2]$ is now the essential spectrum of the bounded operator A_∞ in ℓ^2.

Remark. In [1,2] the authors studied the attracting properties of the spectrum of A_∞ in the case when A_∞ is a compact perturbation of J_∞. The celebrated theorem of H. Weyl claims that $\Sigma(A_\infty) = [-2, 2] \cup \Sigma_d(A_\infty)$, where the discrete spectrum $\Sigma_d(A_\infty)$ is at most denumerable set of eigenvalues $\lambda_j(A_\infty)$ of the finite algebraic multiplicity v_j, off the essential spectrum $[-2, 2]$. It is proved in [1,2] that each λ_j is the attracting point of $\Sigma(A_n)$ of order v_j. Our result in Corollary 3.7 supplements this one nicely. Note that in the case (26) the Weyl theorem is only partly true (see [7, Theorems 7 and 9]): still $[-2, 2] \subset \Sigma(A_\infty)$, but in general there is no discrete part of the spectrum any more.

4. Asymptotically periodic Jacobi matrices: the block case

We start out with the definition of the spectral distribution with matrix-valued symbols. Throughout the rest of the paper θ will stand for a $k \times k$ matrix-valued and the Lebesgue integrable function (i.e., all its entries are integrable) with the eigenvalues $\lambda_j(\theta)$, $j = 1, 2, \ldots, k$.

Definition 4.1. Let θ be a $k \times k$ matrix-valued the Lebesgue integrable function defined on a set G of finite and positive Lebesgue measure. A matrix sequence $\{A_n\}$ has the *asymptotic spectral distribution* θ if for all $F \in \mathcal{C}_0$ one has

$$\lim_{n\to\infty} \Sigma_\lambda(F, A_n) = \frac{1}{km(G)} \sum_{j=1}^{k} \int_G F(\lambda_j(\theta(t)))\, dt.$$

As in the scalar case, we write in short $\{A_n\} \sim_\lambda (\theta, G)$.

Under the essential range of θ we mean now the set

$$S(\theta) := \bigcup_{j=1}^{k} \text{Range}\,(\lambda_j(\theta)).$$

The same argument as applied above in the proof of Theorem 2.4 leads to the following result.

Theorem 4.2. *Let θ be a $k \times k$ matrix-valued Lebesgue integrable function defined on a set G of finite and positive Lebesgue measure, and $S = S(\theta)$ be the essential range of θ. Let $\{A_n\}$ be a matrix sequence distributed as θ in the sense of eigenvalues. Then*

(a) *$S(\theta)$ is a weak cluster for $\{A_n\}$;*
(b) *each point $s \in S(\theta)$ strongly attracts Σ_n with infinite order $r(s) = \infty$.*

Definition 4.3. Let b be a $k \times k$ matrix-valued and the Lebesgue integrable function defined on $[-\pi, \pi)$ with the Fourier coefficients

$$\hat{b}_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} b(t) \exp(-ijt)\, dt \in \mathbb{C}^{k \times k}, \qquad j \in \mathbb{Z}. \tag{29}$$

The function b is called the *generating function* of the sequence of block Toeplitz matrices

$$T_n(b) = \begin{bmatrix} \hat{b}_0 & \hat{b}_{-1} & \cdots & \hat{b}_{1-n} \\ \hat{b}_1 & \hat{b}_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{b}_{-1} \\ \hat{b}_{n-1} & \cdots & \hat{b}_1 & \hat{b}_0 \end{bmatrix} \in \mathbb{C}^{kn \times kn}.$$

It is easy to observe that $T_n(b)$ is Hermitian for every n if and only if its generating function b is Hermitian for almost every $t \in [-\pi, \pi)$, and the index n here denotes the block order.

Let us define a matrix sequence $\{\tilde{T}_m(b)\}$ by the following recipe: $\tilde{T}_{kn} := T_n$, and $\tilde{T}_{kn-j}$ is obtained from T_n by deleting the last j rows and columns for $j = 1, 2, \ldots, k-1$. In other words, $\tilde{T}_m$ is the principal $m \times m$ block of the infinite block-matrix $T_\infty(b) = \{\hat{b}_{p-q}\}_{p,q=0}^{\infty}$.

The following general result due to Tilli (see [19]) is very important in our context.

Theorem 4.4. *If b is any Hermitian-valued and absolutely integrable function on* $[-\pi, \pi]$

$$\int_{-\pi}^{\pi} \|b(t)\|\, dt < +\infty,$$

where $\|\cdot\|$ is any matrix norm in $\mathbb{C}^{k \times k}$, then $\{\tilde{T}_m(b)\} \sim_\lambda$ $(b, [-\pi, \pi])$ in the sense of Definition 4.1.

Now we turn to the case of asymptotically periodic Jacobi matrices. Let

$$J_\infty^{(0)} = \begin{bmatrix} b_0^{(0)} & a_1^{(0)} & & & \\ a_1^{(0)} & b_1^{(0)} & a_2^{(0)} & & \\ & a_2^{(0)} & b_2^{(0)} & a_3^{(0)} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}, \quad a_n^{(0)} > 0, \;\; b_n^{(0)} \in \mathbb{R} \tag{30}$$

be an infinite Jacobi matrix with k-periodic entries

$$a_{n+k}^{(0)} = a_n^{(0)}, \quad b_{n+k}^{(0)} = b_n^{(0)}, \ \ n \in \mathbb{Z}_+, \tag{31}$$

and $\mathbf{a} = (a_0^{(0)}, a_1^{(0)}, \ldots, a_{k-1}^{(0)})$, $\mathbf{b} = (b_0^{(0)}, b_1^{(0)}, \ldots, b_{k-1}^{(0)})$ be two real vectors of order k, which define completely the entries of the whole matrix $J_\infty^{(0)}$. In that case the principal $m \times m$ block of $J_\infty^{(0)}$ in (30) is denoted more explicitly by $J_m^{(0)} = J_m[\mathbf{a}, \mathbf{b}]$.

Let $\theta(\mathbf{a}, \mathbf{b}, t)$ be the Hermitian matrix-valued trigonometric polynomial of the form

$$\theta(\mathbf{a}, \mathbf{b}, t) = J_k[\mathbf{a}, \mathbf{b}] + \begin{bmatrix} 0 & \cdots & 0 & a_0^{(0)} \exp(it) \\ \vdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \vdots \\ a_0^{(0)} \exp(-it) & 0 & \cdots & 0 \end{bmatrix}. \tag{32}$$

It is a matter of simple computation to verify that θ has only three nonzero Fourier coefficients $\hat{\theta}_0$ and $\hat{\theta}_{\pm 1}$,

$$T_n(\theta(\mathbf{a}, \mathbf{b})) = \begin{bmatrix} \hat{\theta}_0 & \hat{\theta}_{-1} & & O \\ \hat{\theta}_1 & \hat{\theta}_0 & \ddots & \\ & \ddots & \ddots & \hat{\theta}_{-1} \\ O & & \hat{\theta}_1 & \hat{\theta}_0 \end{bmatrix} = J_{kn}[\mathbf{a}, \mathbf{b}],$$

and in general $\tilde{T}_m(\theta(\mathbf{a}, \mathbf{b})) = J_m[\mathbf{a}, \mathbf{b}]$ for all m. So the asymptotic distribution for the periodic Jacobi matrix sequence is a particular case of Theorem 4.4 with $b = \theta(\mathbf{a}, \mathbf{b})$. Such asymptotic distribution, paraphrased as the asymptotic distribution of the zeros of orthogonal polynomials p_n (28) with periodic recurrence coefficients, is well known (see, e.g., [21, Section 3] and references therein). The essential range $S(\theta(\mathbf{a}, \mathbf{b}))$ is tightly related to the support of the corresponding orthogonality measure (cf. [12, Theorem 13]). If $k = 1$, then $\theta(\mathbf{a}, \mathbf{b}) = b^{(0)} + 2a^{(0)} \cos t$ and putting $b^{(0)} = 0$ and $a^{(0)} = 1$ we come to the Toeplitz matrix (1).

As in the scalar case ($k = 1$), we are interested in generic complex perturbations of $J_\infty^{(0)}$. An infinite complex Jacobi matrix

$$J_\infty = \begin{bmatrix} b_0 & c_1 & & & \\ a_1 & b_1 & c_2 & & \\ & a_2 & b_2 & c_3 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}, \quad a_n, b_n, c_n \in \mathbb{C} \tag{33}$$

is called the *Cesàro asymptotically k-periodic* if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (|a_j - a_j^{(0)}| + |b_j - b_j^{(0)}| + |c_j - a_j^{(0)}|) = 0,$$

the *asymptotically k-periodic* if

$$\lim_{n\to\infty} (|a_n - a_n^{(0)}| + |b_n - b_n^{(0)}| + |c_n - a_n^{(0)}|) = 0,$$

and the *trace class asymptotically k-periodic* if

$$\limsup_{n\to\infty}\sum_{j=1}^{n}(|a_j-a_j^{(0)}|+|b_j-b_j^{(0)}|+|c_j-a_j^{(0)}|)<\infty,$$

for some k-periodic sequences $\{a_n^{(0)}, b_n^{(0)}\}$ as in (31). In other words, $J_\infty = J_\infty^{(0)} + P_\infty$ with the k-periodic $J_\infty^{(0)}$ (30) (called the background) and the Cesàro compact (compact, the trace class) perturbation P_∞.

The following results can be proved in exactly the same fashion as Corollaries 3.7 and 3.8. In the latter case Theorem 3.6 comes into play. The point is that the essential range $S(\theta(\mathbf{a},\mathbf{b}))$ is now a union of at most k disjoint closed intervals, and all the eigenvalues of $J_n^{(0)}$ (the zeros of orthogonal polynomials $p_n^{(0)}$ (28)), but finitely many (at most $2k$), lie in $S(\theta(\mathbf{a},\mathbf{b}))$. So, in particular, the matrix sequence $\{J_n^{(0)}\}$ is strongly clustered at $S(\theta(\mathbf{a},\mathbf{b}))$.

Theorem 4.5. *Let J_∞ be the Cesàro asymptotically k-periodic Jacobi matrix with the background $J_\infty^{(0)}$ and $\theta(\mathbf{a},\mathbf{b})$ (32) the generating function for $J_\infty^{(0)}$. Then $\{J_n\}$ is distributed as $(\theta(\mathbf{a},\mathbf{b}), [-\pi,\pi])$ in the sense of eigenvalues, weakly clustered at $S(\theta(\mathbf{a},\mathbf{b}))$, and $S(\theta(\mathbf{a},\mathbf{b}))$ strongly attracts the spectra of $\{J_n\}$ with infinite order of attraction for any of its points.*

Theorem 4.6. *Let J_∞ be the trace class asymptotically k-periodic Jacobi matrix with the background $J_\infty^{(0)}$ and $\theta(\mathbf{a},\mathbf{b})$ (32) the generating function for $J_\infty^{(0)}$. Then $\{J_n\}$ is strongly clustered at $S(\theta(\mathbf{a},\mathbf{b}))$, and $S(\theta(\mathbf{a},\mathbf{b}))$ strongly attracts the spectra of $\{J_n\}$ with infinite order of attraction for any of its points.*

4.1. Concluding remarks and further generalizations

As a conclusion, we observe that tools from matrix theory [3,4] combined with those from asymptotic linear algebra [19,20,15] have been crucial for proving plainly results concerning non-Hermitian perturbation of Jacobi matrix sequences. A special part of them is the GLT theory (see [16,17] and references therein) which allows to treat the case of variable coefficients under very mild restrictions on the regularity of the coefficients (e.g., numerical approximations of variable coefficient PDEs [16] and systems of PDEs [17], Jacobi sequences with asymptotically varying periodic [5] and non-periodic [11] coefficients, etc.). The interesting fact is that the tools explicitly developed here are applicable verbatim to these cases as well (see [9] for an example), by allowing to deal with non-Hermitian perturbations under the same mild trace conditions.

Appendix A. Equivalence of trace-norm and entry-wise conditions

Let $A = \{a_{j,k}\}_{j,k=1}^{n}$ be a complex matrix of size n, let $\|\cdot\|_1$ be the trace-norm, and let $\|\cdot\|_{[1]}$ be the componentwise l^1 norm:

$$\|A\|_1=\sum_{j=1}^{n}\sigma_j, \quad \|A\|_{[1]}=\sum_{j,k=1}^{n}|a_{j,k}|$$

with $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$ being the singular values of A. With the notations (i) and (ii) at the end of Section 1, we would like to prove that $\|P_n\|_1 = o(n)$ if and only if (7) holds and $\|P_n\|_1 = O(1)$ if and only if (8) is satisfied. Taking into account the definition of the norm $\|\cdot\|_{[1]}$ and the tridiagonal structure of P_n, $n \geqslant 1$, condition (7) can be rewritten as $\|P_n\|_{[1]} = o(n)$ and, similarly, (8) is equivalent to $\|P_n\|_{[1]} = O(1)$. Therefore, what we would like to prove is the asymptotic equivalence, independently of the size n, of the two norms $\|\cdot\|_1$ and $\|\cdot\|_{[1]}$. Specifically, we look for two positive constants c and C independent of n such that $c\|A\|_1 \leqslant \|A\|_{[1]} \leqslant C\|A\|_1$ for every complex matrix A of size n. For a fixed n, the existence of the two positive constants $c = c(n)$ and $C = C(n)$ is trivial thanks to the topological equivalence of norms in any finite dimensional vector space. The nontrivial part is to show that c and C can be chosen independently of n. Unfortunately, the latter is in general false as the following example shows. Take $A = \{a_{j,k}\}_{j,k=1}^n$ with $a_{j,k} = 1$, $\forall j, k = 1, \ldots, n$. Then $\sigma_1 = n$, $\sigma_2 = \cdots = \sigma_n = 0$, and therefore $\|A\|_1 = n$ while $\|A\|_{[1]} = n^2$ so that $C(n) \geqslant n$ (indeed it can be proved that the previous example is an extremal one and indeed the best constant C is exactly $C(n) = n$).

Therefore, the equivalence of the trace-norm and of the l^1 entry-wise norm has to exploit the fact that the involved matrices are tridiagonal. In the subsequent steps we will use the Fourier analysis of matrices introduced by Bhatia in [4]. Let A be a generic tridiagonal matrix of size n and, for any $m = 1 - n, \ldots, n - 1$, let $\mathcal{D}_m(A)$ be the matrix which coincides with the mth diagonal of A, i.e., $\{\mathcal{D}_m(A)\}_{j,k} = a_{j,k}$ if $j - k = m$ and $\{\mathcal{D}_m(A)\}_{j,k} = 0$ otherwise. Therefore,

$$A = \sum_{m=-1}^{1} \mathcal{D}_m(A) \tag{34}$$

and, by the structure of any $\mathcal{D}_m(A)$, a plain check shows that

$$\|\mathcal{D}_m(A)\|_{[1]} = \|\mathcal{D}_m(A)\|_1. \tag{35}$$

Consequently, by the definition of $\|\cdot\|_{[1]}$, (34), and (35) we have

$$\begin{aligned}\|A\|_1 &= \left\| \sum_{m=-1}^{1} \mathcal{D}_m(A) \right\|_1 \leqslant \sum_{m=-1}^{1} \|\mathcal{D}_m(A)\|_1 \\ &= \sum_{m=-1}^{1} \|\mathcal{D}_m(A)\|_{[1]} = \|A\|_{[1]}\end{aligned}$$

and so $c = 1$ which is independent of n. For the reverse inequality we have

$$\begin{aligned}\|A\|_{[1]} &= \left\| \sum_{m=-1}^{1} \mathcal{D}_m(A) \right\|_{[1]} = \sum_{m=-1}^{1} \|\mathcal{D}_m(A)\|_{[1]} \\ &= \sum_{m=-1}^{1} \|\mathcal{D}_m(A)\|_1 \leqslant 3\|A\|_1,\end{aligned}$$

where for the last inequality we use the identity (see [4])

$$\mathcal{D}_m(A) = \frac{1}{2\pi} \int_{-\pi}^{\pi} D(t) A D^*(t) \exp(-imt)\, dt,$$

with $D(t)$ a diagonal unitary matrix whose jth diagonal entry equals $\exp(i(j-1)t)$. From the latter identity, since the trace-norm is a unitarily invariant norm (see [3]), it easily follows that $\|\mathcal{D}_m(A)\|_1 \leqslant \|A\|_1$. We conclude that $C = 3$ which is again a constant independent of n, as desired.

As already pointed out in the Introduction only the proof is new. The result can be recovered directly from known facts: for instance, use inequalities (2.32) in [10] with $p = 1$ and the (trivial) equivalence between l^∞ and l^1 norms for vectors of size 3. Then one arrives to

$$\frac{1}{3}\|A\|_1 \leqslant \|A\|_{[1]} \leqslant 9\|A\|_1$$

for every tridiagonal matrix A. Note, however, that our constants $c = 1$ and $C = 3$ are tighter and indeed $c = 1$ is optimal (take A the identity matrix).

Finally, it should be remarked that the similar equivalence results can be obtained for more general patterns. Instead of tridiagonal structures we could equally well have considered banded structures (also in a multilevel sense, see [8]). In that case, the proofs are identical and the constants are $c = 1$ and C equals the number of nonzero diagonals of the considered band matrices. As long as this number is independent of n, the two norms $\|\cdot\|_{[1]}$ and $\|\cdot\|_1$ are asymptotically equivalent, i.e., with equivalence constants positive and independent of the size n.

References

[1] D. Barrios, G. López, A. Martínez-Finkelshtein, E. Torrano, On the domain of convergence and poles of complex J-fractions, J. Approx. Theory 93 (1998) 177–200.

[2] D. Barrios, G. López, A. Martínez-Finkelshtein, E. Torrano, The finite dimensional approximation of the resolvent of infinite banded matrix and continuous fractions, Mat. Sb. 190 (1999) 23–42.

[3] R. Bhatia, Matrix Analysis, Springer, New York, 1997.

[4] R. Bhatia, Pinching, trimming, truncating, and averaging of matrices, Amer. Math. Monthly 107 (2000) 602–608.

[5] D. Fasino, S. Serra-Capizzano, From Toeplitz matrix sequences to zero distribution of orthogonal polynomials, Contemp. Math. 323 (2003) 329–340.

[6] J.S. Geronimo, E.M. Harrell II, W. Van Assche, On the asymptotic distribution of eigenvalues of banded matrices, Constr. Approx. 4 (1988) 403–417.

[7] L. Golinskii, On the spectra of infinite Hessenberg and Jacobi matrices, Matematicheskaya Fizika, Analiz, Geometriya 7 (2000) 284–298.

[8] M. Hladnik, J. Holbrook, S. Serra-Capizzano, Fat diagonals and Fourier analysis, SIAM J. Matrix Anal. Appl. 24 (4) (2003) 1060–1070.

[9] S. Holmgren, S. Serra-Capizzano, P. Sundqvist, Can one hear the composition of a drum? Mediterr. J. Math., to appear.

[10] R. Killip, B. Simon, Sum rules for Jacobi matrices and their applications to spectral theory, Ann. Math. 158 (2003) 253–321.

[11] A.B.J. Kuijlaars, S. Serra-Capizzano, Asymptotic zero distribution of orthogonal polynomials with discontinuously varying recurrence coefficients, J. Approx. Theory 113 (2001) 142–155.

[12] A. Máté, P. Nevai, W. Van Assche, The support of measures associated with orthogonal polynomials and the spectra of related self-adjoint operators, Rocky Mountain J. Math. 21 (1991) 501–527.

[13] P. Nevai, Géza Freud, orthogonal polynomials and Christoffel functions, A case study, J. Approx. Theory 48 (1986) 3–167.

[14] W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1974.

[15] S. Serra-Capizzano, Spectral behavior of matrix sequences and discretized boundary value problems, Linear Algebra Appl. 337 (2001) 37–78.

[16] S. Serra-Capizzano, Generalized locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations, Linear Algebra Appl. 366 (1) (2003) 371–402.

[17] S. Serra-Capizzano, The GLT class as a Generalized Fourier Analysis and applications, Technical Report, SCCM-05-07, Stanford University 2005, Linear Algebra Appl., to appear.

[18] H. Stahl, V. Totik, General orthogonal polynomials, Encyclopedia of Mathematics and its Applications, vol. 43, Cambridge University Press, New York, 1992.

[19] P. Tilli, A note on the spectral distribution of Toeplitz matrices, Linear Multilinear Algebra 45 (1998) 147–159.

[20] P. Tilli, Some results on complex Toeplitz eigenvalues, J. Math. Anal. Appl. 239 (2) (1999) 390–401.

[21] W. Van Assche, Zero distribution of orthogonal polynomials with asymptotically periodic varying recurrence coefficients, in: V.B. Priezzhev, V.P. Spiridonov (Eds.), Self-Similar Systems, Joint Institute for Nuclear Research, Dubna, Russia, 1999, pp. 392–402.